NYSABE Journal, 11, 1996

Effect of Open-Ended Questions on Spanish-Dominant LEP Students' Ability to Demonstrate In-Depth Science Concept Development and Use Scientific Vocabulary

María Aguirre
SUNY College at Buffalo

This study examined the effect of two on-process alternative assessment modes on Spanish-dominant (LEP) students' performance on in-depth concept development and scientific vocabulary use. Forty three low and middle socioeconomic status fifth grade LEP students and their respective teachers participated in this study. Below-average and above-average science achievement boys and girls were randomly assigned to two groups. Students in both groups were asked to answer the same open-ended question at three assessment times during the study of a chapter on electricity and magnetism. Students in group one were asked to answer the question by writing a paragraph and students in group two answered the same question by making a drawing. Results from a repeated measures MANOVA suggest that, in general, the open-ended drawing mode seemed to be an appropriate assessment mode particularly for below-average LEP students. Results from unplanned ANOVA's revealed that the drawing mode beneficiated more the boys than the girls across time. The students expressed preference for the drawing mode of assessment because they felt more confident and comfortable answering questions by drawing.

Introduction

Purpose of the Study

The purpose of this study was to examine the effect of two on-process alternative assessment modes on students' performance on in-depth concept development and scientific vocabulary use. The modes were open-ended questions answered by writing a paragraph (OE-W) and by drawing (OE-D). Students' opinions and feelings about the assessment modes were also explored. Participants were fifth grade Spanish-dominant LEP students at various levels of science achievement, taught in their regular classroom settings.

Review Of Related Literature

Current science curriculum and assessment reform movements across the nation are primarily based on constructivist theory (High School Framework for National Science Education Standards, l995; Developing Biological Literacy, l995; and Benchmarks for Science Literacy, Project 2061, l993).

According to this theory, learning involves a process by which learners can construct experiences in order to develop concepts or mental schemata. Kibby (l995) states that concepts are mental structures learners have in their minds to represent objects, actions, ideas, and feelings. Eeds & Cockrum (l985) and Saunders (l992) posit that schemata are structures that represent the network of interrelations believed to normally hold among the subconcepts or constituents of a concept.

There are many aspects influencing each human being's concept construction every child in a classroom can have a different interpretation of a concept taught. Human thinking always takes place in some context and is influenced by it. People have thinking skills at different levels of development, have different levels of imagination, and are always engaging in metacognition. Students' interpretations are unique and most of them might be acceptable (Department of Public Education of Puerto Rico, l987). However, most traditional tests ask students to choose one correct phrase or a word which represents the teacher's "truth". There is a great chance that the teacher's "truth" may not correspond to the student's mental schemata. In this case, the teacher can wrongly assume that the student did not master the concept under consideration, when in fact, the student could have some degree of understanding or even very sophisticated ideas about the assessed concept. In contrast, when the students are given the opportunity to show their unique understanding of the concept, as it happens in on-process open-ended questions, where more than one truth is accepted, the students have better chances to demonstrate their developing knowledge. This, in turn, might improve students' attitude and disposition for learning. Moreover, students' responses may help teachers to revise their teaching-assessment methods for the benefit of the students.

Many of the previously mentioned curriculum reform programs in science education structure their framework in accordance with the specifications of the National Science Education Standards Document (National Research Council, l996). Assessment Standard D in the NRC Document states that "assessment tasks must be modified appropriately to accommodate the needs of students with physical disabilities, learning disabilities or limited English proficiency".

In response to Assessment Standard D and similar important standards included in the National Research Council (l996); the National Science Teachers Association (l995), the Biological Science Curriculum Study (l995), and the American Association for the Advancement of Science (l993) recommend the use of alternative assessment techniques to assess students' learning processes. Alternative assessment refers to the process by which students can show the quantity and quality of their learning processes in ways that best fit their individual characteristics. It allows students to show what they can do with the knowledge they have learned at different times and in different contexts. Alternative assessment implies that each assessment activity during the process should be an additional opportunity for students to learn and for teachers to assess them (Meyer, 1992; Aguirre, l992, l993; Marzano, Pickering and McTighe, l993).

These frameworks advocate modes of alternative and authentic assessment techniques that help students to develop concepts that they can "file" in their long term memory (Doran, Tamir and Chan, l994; Puerto Rico Statewide Systemic Initiative, l994). Some of the alternative assessment techniques mentioned are portfolios, performance tasks, and higher level thinking pencil and paper tasks. However, no empirical data are presented in any of the previously mentioned science reform frameworks to support the effectiveness of these techniques (Carlson, l991; Smith and Welliver, l990). Gil-Pérez and Carrascosa (l990), Dreyfus, Jungwirth and Eliovitch (l990), and Finley, Lawrenz and Heller (l992) argue that research in science education should pay more attention to alternative assessment and their implications for the education of minorities. This study addresses this issue by examining the use of two alternative assessment modes, open-ended drawing and open-ended writing with Hispanic-Spanish dominant students.

Definitions

For this study, on-process open-ended questions were questions open enough to include the content of a whole chapter in electricity and magnetism. Two modes of open-ended questions were used: drawing and writing. Students answered versions of the open-ended questions at different times during the study of a chapter. Students' answers were expected to illustrate if students' schemata changed as a result of the interaction of prior schemata with learnable content. By monitoring students' learning at different times, it was expected that teachers and students could become more aware of the quality of their learning. According to Marzano, et al. (l993) going over their learning processes permits the teacher and the student to determine what the student learned, and what the student needs to learn, and also allows for the correction of any misconception.

Students' performance on in-depth concept development meant that students could: 1) illustrate attributes of the concepts, 2) make connections and interconnections between and among subconcepts that were related to the learnable content, and/or illustrate connections with concepts already learned from prior topics (chapters), and 3) write or draw their own interpretations of the concepts.

Students' use of scientific vocabulary represented the use of scientific words included in the selected science content. According to Lee and Fradd (l995), science vocabulary as well as science knowledge, i.e., what students know about science, are important elements in developing science literacy, particularly for students who must learn the academic language of science while developing proficiency in English.

Research Questions

This study attempted to examine the use of OE-D and OE-W assessment modes by Spanish-dominant LEP students to demonstrate concept development and use of scientific vocabulary.

The following questions guided the study: 1) Are there significant differences in the students' performance on in-depth concept development (IDC) and use of scientific vocabulary (VOC) across three assessment times when the students' answers to open-ended questions using drawings (OE-D) are compared with the students' answers to the same questions using writing paragraphs (OE-W)?, 2) Are there significant differences in IDC and VOC among students classified as below or above average in their science achievement when their responses to OE-D or OE-W across the three assessment times are considered?, and 3) What are the students' feelings and attitudes with respect to OE-D and OE-W assessment modes?

Method

Pilot Study

A pilot study was conducted to test and validate the two alternative assessment modes, OE-D and OE-W. This pilot study was conducted during March and April of l996 in order to create and test the instruments and procedures used in the reported here study. A sample of 21 Spanish-dominant LEP students and their teacher participated in the pilot study.

Instruments

Seven instruments were developed during the Pilot Study. They were also tested for reliability and validity.

Instrument 1. Open-ended Questions. For the experimental part of this study, six versions of the open-ended question created during the Pilot Study were used. One open-ended question was developed for each assessment time (prior to the beginning, in the middle and at the end of the chapter) in each of the two assessment modes (OE-D and OE-W). An English translation of the questions based on the topics Electricity and Magnetism follows: 1. Assessment Time: Prior to beginning of the chapter. The students received the following task:

a. Group one (OE-W): Imagine that you are a teacher and you are going to explain to the students all you know about Electricity and Magnetism. In the space below, write a paragraph including all the scientific ideas that you are going to teach and how you are going to explain your ideas.

b. Group two (OE-D): Imagine that you are a teacher and you are going to explain to the students all you know about Electricity and Magnetism. In the space below, draw a picture including all the scientific ideas that you are going to teach. You can use words, phrases, arrows, and other symbols to complete or connect your ideas in order to show how you are going to explain your ideas.

2. Assessment Time: In the middle and at the end of the chapter. During the second and third assessment times the students were asked to go over their previous paragraphs or drawings to correct, amplify their ideas, and present versions to illustrate what they had learned including examples not studied in class.

Instrument 2. Guide for the Administration of the Assessment. This guide described the steps the teachers needed to follow to assure consistency in the administration of the two assessment modes in both classrooms.

Instrument 3. Rubric. This instrument was designed for the correction of students' responses to the open ended questions in both assessment modes. The rubric had the following six levels of performance: Level 5 - Excellent answer (22-28 points); Level 4 - Competent answer (16-21 points); Level 3 - Answer with minor errors but satisfactorily (9-15 points); Level 2 - Starts but fails to complete the task (3-8 points); Level 1 - Incapable of an effective beginning (2 points), and Level 0 - There is NO attempt to do the task (0 points).

The rubric was modified for the study reported based on: 1) the chapter objectives, 2) the list of scientific vocabulary and 3) the concept map created by the researcher to show the connections and interconnections among the scientific terms. The inter-rater reliability of the rubric was established. The correlations between scorers and the researcher were: .85, .92, and .92, respectively.

Instrument 4. Scoring Checklist. This checklist permitted raters to register the levels of performance and the points the students received in each criteria for each assessment time. Raters could make comments justifying the points awarded and the level of performance assigned.

Instrument 5. Guide for the Scoring Process. This guide was an adaptation of the procedure that the California Assessment Program and the Puerto Rico State Wide Systemic Initiative Assessment Program follow to score the state assessment tests. It was also based on the Guidelines for Developing and Scoring Free Response Test (Educational Testing Service, l987; Puerto Rico Statewide Systemic Initiative, l994). The guide defines the steps that the scoring committee members were required to follow during the calibration and scoring procedures to establish reliability. The main adaptations were: a) the use of a system of points to score the students' responses in addition to providing the levels of performance, b) the use of cumulative total points for the students' responses at each assessment time, i.e., the score a student obtained during the first assessment was added to the score obtained during the second assessment, etc.

Instrument 6. Guide for Student Interviews. This guide consisted of various questions designed to help students express: a) their understanding of the concepts learned, b) science vocabulary acquired, and c) the strategy used to answer the questions during each assessment.

Instrument 7. List of the Scientific Words. This list contained 23 concepts embedded in the selected chapter. The list was used to determine the scientific vocabulary students used in their responses.

Procedures

Subjects

Forty three 10 to 11 years old fifth grade students of low and middle socioeconomic status in an urban school district in Western New York participated in the study. The subjects were 23 girls and 20 boys, all native speakers of Spanish, enrolled in two fifth grade bilingual science classrooms. Students were classified as LEP with a score of five or less in the Language Assessment Skills Test (LAS, Duncan and De Ávila, l990). English and Spanish versions of the LAS were administered in 1995. Thirty students scored three or below in the English version of LAS. Thirteen subjects scored 4 or 5 in both English and Spanish versions of the LAS. However, these thirteen subjects also scored below the 40th percentile on the CTBS English Reading Test. Therefore, all forty-three subjects were identified as Spanish-dominant LEP students.

Students science achievement was classified as below-average and above-average according to students' scores on the district Examination: Science - Grade Fourth - June (Buffalo Public Schools, Department of Science Education, l995). The test was administered when the participant students were in fourth grade. The distribution of science achievement scores into below and above average was drawn by determining the central score that included equal amount of scores in both groups. The distribution was as follows: below average = 70-75 and above average = 76-97.

At the beginning of the study there were twenty five students in one classroom and twenty two in the other. In order to decrease threats to the internal and external validity due to teacher differences, half of the students in each class were randomly assigned to group one (OE-W) and the other half to group two (OE-D). This way, group one and group two were composed of students from both classes and included boys and girls of varying science achievement.

Because of student transfers or illnesses a total of four children, two girls and two boys, were not available to complete the on-process assessment. Three of these students belonged to group one and one to group two. All four children were classified as below average. The number of participanting students was, therefore, reduced to forty three students.

The participant teachers regularly taught the bilingual science classes. Both teachers hold a New York State Certification in K-6 grade and in bilingual education, and have two or more years of experience teaching science to LEP students in this school system.

Procedures Prior to the Study

Teachers' Training and Conferences

The researcher provided training to the teachers on how to implement the alternative assessment modes used in this study. The researcher and the teachers met to discuss the content, make possible adaptations to the chapter, and discuss the teaching strategies to use during the instruction of the selected chapter.

Students' Training

The teachers provided training to students in order to familiarize them with the assessment technique. Each teacher conducted activities in which the students were asked to answer open-ended questions in the modality that they would be assigned later on (OE-D or OE-W).

Procedures During the Study

Teaching

The two teachers were guided to teach the science class in Spanish using student centered teaching techniques. The science content used in this study belongs to an existing elementary science curriculum (Science Horizons, chapter 8, Mallinson, Mallinson, Froschauer, Harris, Lewis & Valentino, l991) which correlates with the District Scope of the Curriculum of the Level III - Grade 5 Syllabus (District Board of Education, l991). Chapter eight in the Science Horizons textbook consisted of four sections: 1) Electric Charges in Atoms, 2) Charged Objects, 3) Moving Electrons, and 4) Electricity and Magnetism. Each teacher administered the same tasks: open-ended questions using OE-W and OE-D modes.

Administration of the Assessment

The administration of the assessment tasks took place at three assessment times during the lessons: (a) prior to the beginning of the chapter, (b) in the middle, and (c) at the end of the chapter. Students' responses prior to the beginning of the chapter represented the students' previous knowledge, i.e., the base upon which the students' future concept and vocabulary development were built. The second assessment was conducted when the first two topics in the chapter were covered. The third assessment took place after a short review of the chapter when all four topics had been covered. During each assessment both groups of students had forty five minutes to answer the questions.

Students in group one answered the questions by writing a paragraph using the scientific vocabulary studied to describe their understanding of the concepts. Students in group two answered the questions by creating a drawing representing their ideas. They were also instructed that the drawing could include words, phrases and other symbols. For example, when the students were asked to represent what they knew about Electricity and Magnetism, they created their own responses using drawings. Symbols (eg., arrows, + or - , etc.) and words were then used to identify things in their drawings. Using this assessment mode, students had to represent not only the attributes of the concepts they had learned but also they had to show the organization of the subconcepts in the situation chosen by them, and the interrelationships among the subconcepts, including those already studied in other chapters.

The researcher observed the process of the task administration at all assessment times. Each assessment time was scheduled at different days, but during the last two morning class periods.

For the second and third assessments, students in both groups received the copies of the responses made during the previous assessment. Students had the opportunity to review their previous responses and correct or modify them, according to what they were learning at the assessment time.

Students' Interviews

All the students also participated in five to ten minute individual interviews with the researcher after each assessment time. In this session, students had the opportunity to reflect on their responses and explain the process they followed in answering the questions. They also talked about the connections they could identify between the concepts. In addition, students could talk about any ideas or concepts they forgot to express in their answers. The reflection processes students underwent during the conferences were tape-recorded. Each conference was analyzed by the researcher after each assessment time. Students' "misconceptions" or incomplete ideas about electricity and magnetism were identified through their responses to the OE-W and OE-D. The ideas identified were compared with what was said during the interviews. A list of the most relevant students' misconceptions were given to the teachers at the first two assessment times in order for the teachers to address students' problems during the following class sessions. The data gathered during the students' interviews were another source to count the scientific vocabulary each student used.

Scoring Process

Three experts, one science coordinator and two fifth grade classroom teachers, constituted the scoring committee. The members of the scoring committee were selected based on their content mastery and availability.

Students responses to the open-ended questions in both groups were scored using the same rubric. During the scoring process all scorers marked the students' answers. Students responses were divided into sets of five for each assessment mode. After each set was corrected, a discussion was held to check the levels and scores given to each student's answer. When the differences for a student response were more than one level and/or when the difference in scores was more than five points, the scorers assigned a final score by consensus.

Measurement of the Dependent Variables

There were two dependent variables: 1) the students' in-depth concept development (IDC) and 2) the students' use of scientific vocabulary (VOC). For IDC, students' scores (points) based on answers to the OE-D or OE-W were used.

For VOC, the investigator counted the selected scientific vocabulary each student used at each assessment time. The researcher selected the last two assessments to compare two methods of counting scientific vocabulary: 1) listening to the tapes recorded after the assessments and 2) examining the students' responses to both modes of open-ended questions. Since there was a moderately-high correlation in times two and three (0.78 and 0.84, respectively) between the two methods of counting, the researcher decided to measure VOC from students' responses to the open-ended questions (drawings or paragraph) which was less time consuming.

Results

Data Analysis

The analysis of the data involved a repeated measures 2X2X3 MANOVA design. Grouping variables included assessment modes (i.e., drawing and writing) and students' science achievement (i.e., above-average and below-average), with assessment time (i.e., prior to the beginning of the chapter, in the middle, and at the end of the chapter) as the within-subjects factor. A polynomial contrast was used for the dependent variable transformation.

Findings

In relation to question one, the mode of assessment X assessment time interaction reached statistical significance, *F* (4,33) = 5.28, *p* = 0.002 (see Table 1). Inspection of the means revealed that students performed better on the drawing mode in Time One and Time Three. A significant effect was also found for mode of assessment, *F* 2,35 = 7.30, *p* = 0.002. Inspection of the IDC means for mode of assessment revealed that students in the drawing group performed better than those in the writing group (see Table 2). Inspection of the VOC means revealed that students in both modes of assessment performed similarly (OE-D = 6.6; OE-W = 6.6).

Effect	F	Effect df	Error	Sig of F
Science Achievement	2.77	2	35	0.076
x Mode of Achievement	5.64	2	35	0.008*
Science Achievement	7.30	2	35	0.002*
Mode of Assessment	2.41	4	33	0.069
Sicence Achievement x Mode of Ass. x Time	1.54	4	33	0.213
Science Achievement x Time	5.28	4	33	0.002*
Time	66.66	4	33	0.000*

Table 1. Manova Results (significant at .05 level)

In-depth Concept Development (IDC)

Mode	Sciece Achievement	Time 1	Time 2	Time 3
Drawing	Above Average	5.7	11.0	17.0
Drawing	Below Average	4.4	9.5	16.5
Writing	Above Average	5.8	10.4	16.2
Writing	Below Average	2.8	5.2	7.9
Scientific Vocabulary Used (VOC)				
Drawing	Above Average	3.2	7.1	11.7
Drawing	Below Average	2.7	5.6	8.8
Writing	Above Average	3.3	8.4	13.3
Writing	Below Average	2.7	4.8	7.2

Table 2. Means: Mode Of Assessment x Science Achievement Across Time

In-depth Concept Development (IDC)

In relation to question two, the interaction of students' science achievement (categorized as below and above average) X assessment time was not statistically significant. Likewise, the three-way interaction of assessment mode X students' science achievement X assessment time was not significant (see Table 1). There was a significant difference between above and below average students, $F(2,35) = 5.64$, $p = 0.008$. It was also found a significant difference related to Time, $F(4,33) = 66.66$, $p = 0.000$. Inspection of the means revealed that for IDC above average students in the drawing group performed better than those in the writing group in assessments two and three (see Table 2). The means of below average students group in the drawing mode were higher than those in the writing mode across time. In relation to VOC, above average students in the writing mode performed better than those in the drawing mode across time. Below average students in the drawing mode performed better than those in the writing mode in assessments two and three.

Question three explored students' opinions and feelings about the assessment modes. The data collected during the student interviews revealed several interesting findings. First, ten students out of 43 (23%) felt confused and anxious while answering the practice questions (i.e., questions similar to those they would answer during actual assessment). After practice all said they gained confidence and were not confused or anxious. Second, all students indicated that it was the first time they had answered these kind of comprehensive open-ended questions (those that covered the entire chapter). They also indicated that it was the first time that they had the opportunity to revise their responses during and after the test. Third, twenty-two of the twenty-three (96%) students who answered questions by drawing felt that answering by drawing was easier than answering by writin. Fourth, two out of ten (20%) above-average students in the OE-W group said that they preferred to answer questions by writing because they enjoyed writing, often writing stories as a hobby. All below average students said drawing could be easier than writing paragraphs, Last, all of the OE-D students said that they enjoyed answering questions by drawing, and that they preferred this kind of assessment, because they had the opportunity to think and use their creativity while answering questions.

Effect of Gender on IDC and VOC Across Time

Six unplanned ANOVAs were conducted to explore the effect of gender for both IDC and VOC for each assessment time (Table 3).

For IDC, there was a significant interaction between gender and mode of assessment for time three, $F(1,39) = 6.58$, $p = 0.015$, but not for time two and time one (Table 3). However, males outperformed females across time on the drawing mode and females outperformed males across time on the writing mode (Table 4).

Using ANOVAs for IDC, significant differences were again found for mode of assessment for time two, $F\ (1,39) = 4.83$, $p = 0.035$ and for time three, $F\ (1,39) = 10.95$, $p = 0.002$. However, mode of assessment for time one failed to reach statistical significance (see Table 3). Nevertheless, when gender was considered, females obtained similar scores on both drawing and writing across time but males obtained higher scores on the drawing mode. Females obtained higher scores for time two and three on the writing mode, while males obtained higher scores for time two and three on the drawing mode (Table 4).

For VOC, interactions were statistically significant between drawing and writing for males and females for time two, $F\ (1,39) = 10.45$, $p = 0.003$ and time three, $F\ (1,39) = 9.40$, $p = 0.004$. However, no significant interaction was found between drawing and writing modes for males and females for time one (Table 3). Females obtained higher scores on the writing mode at all assessment times. Males obtained higher scores for all assessment times on the drawing mode. No statistically significant differences were observed between mode of assessment (writing and drawing) and VOC during any of the three assessment times.

Discussion

In general, students who answered the open-ended questions by

Dependent Variable	Source	df	E	F	Sig of F
IDC[1] - T_1	Mode	1	39	1.09	0.303
	Sex	1	39	0.00	0.980
	Mode x Sex	1	39	0.86	0.361
IDC[1] - T_2	Mode	1	39	4.83	0.035*
	Sex	1	39	0.42	0.521
	Mode x Sex	1	39	3.90	0.056
IDC[1] - T_3	Mode	1	39	10.95	0.002*
	Sex	1	39	0.54	0.469
	Mode x Sex	1	39	6.58	0.015*
VOC[2] - T_1	Mode	1	39	0.02	0.898
	Sex	1	39	0.11	0.743
	Mode x Sex	1	39	3.01	0.091
VOC[2] - T_2	Mode	1	39	0.00	0.975
	Sex	1	39	0.00	0.975
	Mode x Sex	1	39	10.45	0.003*
VOC[2] - T_3	Mode	1	39	0.06	0.816
	Sex	1	39	0.01	0.947
	Mode x Sex	1	39	9.40	0.004*

[1] In-depth Concept Development [2] Science Vocabulary Used
* Significant at the .05 level

Table 3. Results of Two-Way ANOVAs: Mode of Assessment X Sex Across Time

Mode	Sex	Time 1	Time 2	Time 3
In-depth Concept Development (IDC)				
Drawing	Male	5.4	11.0	18.2
Drawing	Female	4.6	9.4	15.3
Writing	Male	3.9	5.9	8.9
Writing	Female	4.5	9.1	14.1
Scientific Vocabulary Used (VOC)				
Drawing	Male	3.3	7.9	12.5
Drawing	Female	2.6	4.6	7.6
Writing	Male	2.4	4.6	7.1
Writing	Female	3.4	7.9	12.3

Table 4. Means: Mode of Assessment X Sex Across Time

drawing demonstrated well their interpretation of science concepts or ideas embedded in the chapter. Students performed better using the drawing mode than when using the writing mode in IDC. The drawing mode permitted students to illustrate better their understanding of their schemata. Drawing also made it easier to detect students' misconceptions such as: 1) "los protones tienen carga negativa" (protons have negative charge), and 2) static electricity is produced from a continuous electron flow. For VOC, students perform equally well in the drawing and the writing modes across time. This finding suggests that the OE-D can also be useful in assessing students' vocabulary.The study also addressed the interaction between assessment mode and science achievement levels. In order to understand the interaction, the means were plotted across time for all combinations of the grouping variables for both dependent measures, IDC and VOC (Figures 1 and 2).

Examination of the plots revealed that with respect to in-depth concept development the students in the above-average science achievement group performed better than those in the below-average group at times two and three. However, the above-average students performed better in the writing mode than in the drawing mode for vocabulary use at times two and three (see Figure 2). The below-average science achievement students performed better in the drawing mode than the writing mode for IDC and for VOC across time.

The above-average students performed considerably better in vocabulary using writing than in vocabulary using drawing. It is assumed that a high score in the science achievement test reflects that not only the student had mastered science knowledge but in addition possessed an extensive science vocabulary. It is possible that having more vocabulary permitted the above-average students, in both assessment modes, better express their ideas and use more scientific vocabulary than the below average students in the writing mode. Below-average students performed poorly in the writing mode probably because the task is highly demanding in terms of vocabulary use. These students probably did not have enough science vocabulary to express their ideas, a notion supported by their poor performance in the science achievement test, an instrument highly dependent in vocabulary knowledge.

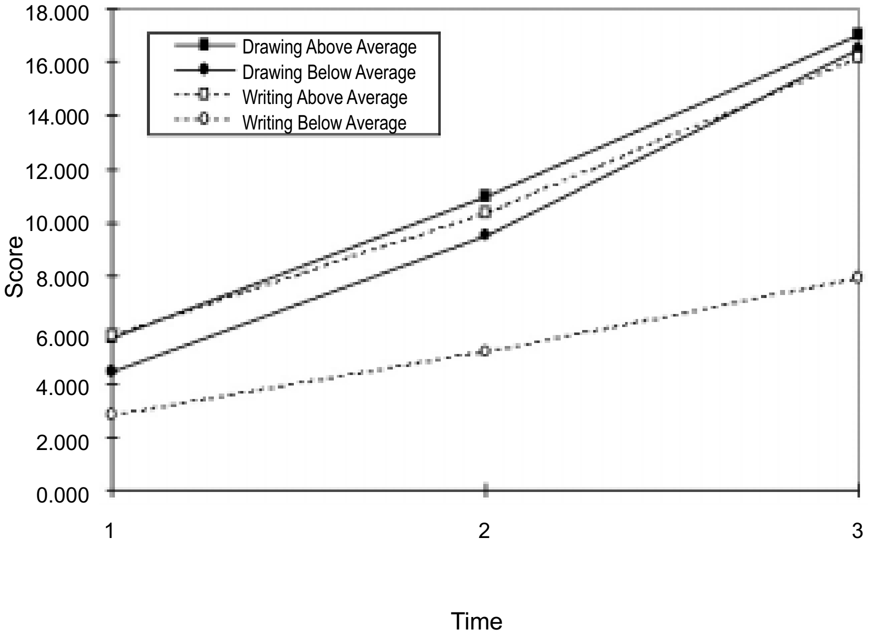

Figure 1. IDC Scores: Plots of Means across Time

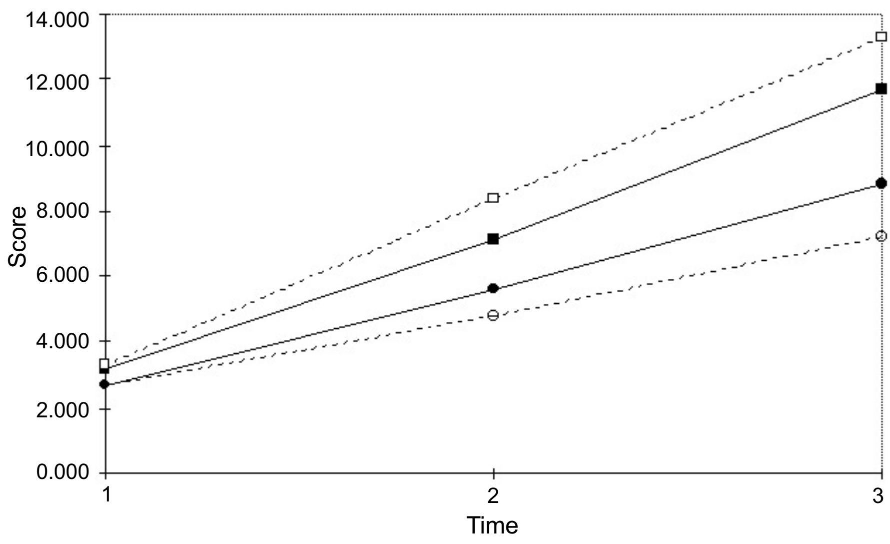

Figure 2. VOC Scores: Plots of Means across Time

Below-average students performed as well as above-average in IDC when using the drawing mode. Below-average students scores (low VOC mean) were not congruent with their in-depth concept development scores (high IDC mean). It seems that in general students' scientific vocabulary (VOC) scores did not necessarily reflect students' in-depth concept development (IDC). For example, many below-average students failed to include words in the drawing mode to identify the scientific concepts they included in their answers but they could demonstrate their understanding of the concepts learned. They used symbols and arrows to illustrate connections among the learned concepts. This explanation is supported by Lee and Fradd (l995) who suggest that when communicating science, the presence or absence of vocabulary does not necessarily reflect students' levels of understanding.

The three-way interaction of assessment mode, science achievement and time suggests that the time-series trend for below-average students is particularly strong when the OE-D mode of assessment is used (Figures 1 and 2). This interaction, however, only approaches statistical significance most likely due to a lack of statistical power (Table 1). Increasing the number of cases from 43 to a larger number would probably produce sufficient power to render the three-way interaction significant.

With respect to time as a factor, participants showed improvement, regardless of assessment mode or their science achievement level specially from time two to time three for IDC in the drawing mode. It appears that students in the drawing mode became more confident in drawing after the second assessment. They overtly expressed this in the interviews. The results suggest that, although the OE-D, in general yielded better scores, both on-process open ended questions could be used to explore Spanish-dominant LEP students' in-depth concept development and scientific vocabulary improvement.

With regard to the effect of gender on IDC and VOC across time, inspection of the plots revealed that for both in-depth concept development and vocabulary the drawing mode favored males (Figures 3 and 4). For the writing mode, male subjects used less scientific vocabulary and showed less understanding of concepts than

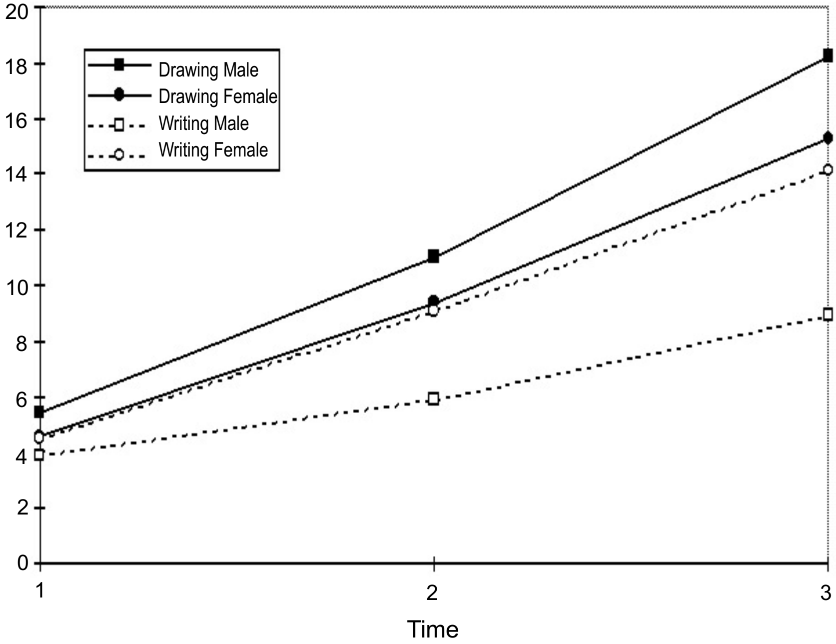

Figure 3. IDC Scores: Means for Mode of Assessment x Sex across Time

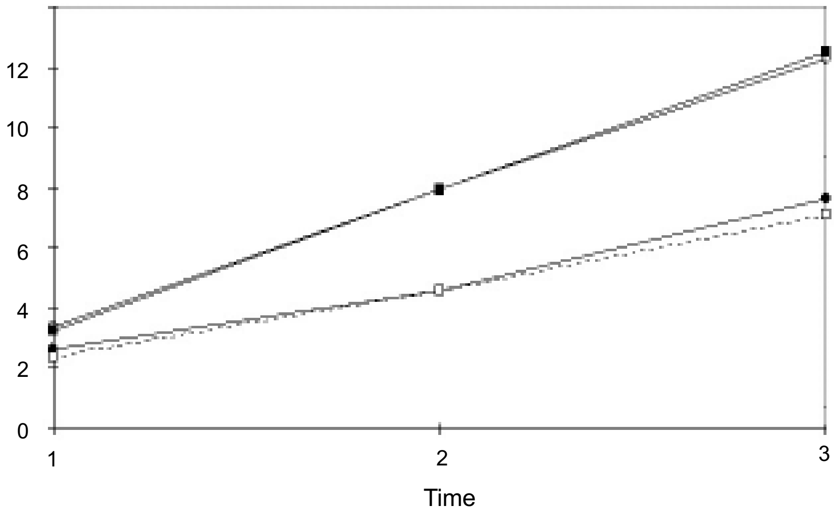

Figure 4: VOC Scores: Means for Mode of Assessment x Sex across Time

females. Males displayed a greater understanding of concepts through drawing than did females. Females however, communicated their knowledge better through writing, but not necessarily by using more science vocabulary than males. For example, the vocabulary mean for males in the drawing mode (Time 3 = 12.55) was close to the mean for females in the writing mode (Time 3 = 12.27). However, females in the writing mode showed lower IDC mean than males did in the drawing mode. Lee and Fradd (l995) reported similar results.They indicated that compared to boys, girls tended to use more language in both oral and written communication.

The third question gathered students' opinions and feelings with respect to the OE-W and OE-D assessment modes. The findings indicated that in general the students preferred OE-D because they felt more confident and comfortable answering questions by drawing.

Rakow & Bermúdez (l993) and Hodson (l993) point out that LEP students have limited communication skills, which do not permit them to express ideas through writing either in English or their native language. This study supports these findings by suggesting that in both modes of assessment students had numerous language problems while expressing ideas in Spanish. Although most language mistakes occurred in spelling, this problem can affect the evaluation of the students' knowledge if these mistakes are interpreted as science misconceptions.

Implications

While it is always difficult to make specific assessment recommendations on the basis of a single study, the findings in general support the use of the drawing mode as an alternative assessment technique with Spanish-dominant LEP students of different science achievement levels.

The information gathered from student responses from both modes can be used by teachers to better direct their teaching in order to address student needs and to help students in concept development. On-process OE-D and OE-W can be alternated with other traditional and alternative modes of assessment. This is one possible response to the claim of national curriculum reform movements that recommend varied modes of assessment to meet the needs of diverse populations.

The preliminary results reported here suggest the need to replicate this study with a larger sample, allowing for a more adequate assessment of three-way interaction of assessment mode X science achievement level X time. Furthermore, using a larger sample will provide a more detailed insight of the different trends between above and below-average science achievement students, especially with respect to scientific vocabulary. There is also a need for a qualitative analysis that would address the issue of how students transform ideas or concepts in relation to electricity and magnetism from one assessment to another.

References

Aguirre, M. (l993). *Desarrollemos nuestra capacidad para pensar--guía del maestro*. San Juan, Puerto Rico: Publicaciones Yuquiyú.

Aguirre, M. (l992). El portafolio: Técnica que promueve la acción, creación y la autoreflexión en el proceso de aprendizaje. *Acta Científica, 6 (1-3)*, 147-150.

American Association for the Advancement of Science (l993). *Benchmarks for science literacy.* New York: Oxford University Press.

Biological Science Curriculum Study (l995). *Developing biological literacy* (ISBN 0-7872-0193-6). Iowa: Kenadall/Hunt.

Buffalo Public Schools, Department of Science Education (l995). *Examination: Science - Grade Four - June, l995*. Buffalo, NY.

Buffalo Board of Education (l991). *Elementary Science Scope of the Curriculum - Level III - Grade 6.*

Carlson, D. (l991). Changing the face of testing in California. *California Curriculum News Report, 16 (3)*, 2-8.

Department of Public Education of Puerto Rico (l987). *Principios para la integración del currículo [Principles for an integrated curriculum]* (DIP No. T321.87). San Juan, Puerto Rico: Taller de Artes Gráficas.

Doran, R. L., Tamir, P., & Chan, A. (l994). *Assessment in Science.* National Science Teachers Association. Unpublished manuscript.

Dreyfus, A., Jungwirth, E., & Eliovitch, R. (l990). Applying the "cognitive conflict" strategy for conceptual change - some implications, difficulties, and problems. *Science Education, 74(5),* 555-569.

Duncan, S. E., & De Ávila, E. A. (l990). *Language Assessment Skills*. Monterrey, Ca.: CTB & McGraw-Hill.

Educational Testing Service (l987). *Guidelines for developing and scoring free-responses tests*. New York: ETS.

Eeds, M., & Cockrum, W. A. (l985). Teaching word meanings by expanding schemata vs dictionary work vs reading in context. *Journal of Reading, 28(6*), 492-497.

Finley, F., Lawrenz, F., & Heller,P. (l992). A summary of research in science education - l990. *Science Education, 76(3),* 239-254.

Gil-Pérez, D., & Carrascosa, J. (l990). What to do about science "misconceptions". *Science Education, 74(5),* 531-540.

Hodson, D. (l993). In search of a rationale for multicultural science education. *Science Education, 77(6),* 685-711.

Kibby, M. W. (l995). The organization and teaching of things and the words that signify them. *Journal of Adolescent and Adult Literacy, 39(3),* 208-223.

Lee, O., & Fradd, S. H. (l995). Science knowledge and cognitive strategy use among culturally and linguistically diverse students. *Journal of Research in Science Teaching 32(8),* 797-816.

Mallinson, G., Mallinson J., Froschauer, L, Harris, J., Lewis, M, & Valentino, C. (1991) *Science Horizons 5*. New Jersey: Silver Burdett & Ginn.

Marzano, R., Pickering, D., & McTighe, J. (l993). *Assessing student outcomes*. Virginia: Association for Supervision and Curriculum Development.

Meyer, C. A. (l992). What's the difference between authentic and performance assessment? *Educational Leadership, 49(8),* 39-40.

National Science Teachers Association. (l995). *A high school framework for national science education standards* (PBI22X).

Virginia: Project on Scope, Sequence, and Coordination of Secondary School Science.

National Research Council (l996). *National science education standards* (ISBN 0-309-05326-9. Washington, DC: National Academy Press.

Puerto Rico Statewide Systemic Initiative. (1994). *Iniciativa sistémica para la excelencia en ciencias y matemáticas. [Statewide systemic initiative for the excellence in Science and Mathematics].* San Juan, PR: Recourse Center for Science and Engineer, UPR.
Rakow, S. J. & Bermúdez, A. B. (l993). *Science Education, 77(6),* 669-683.

Saunders, W. L. (l992). The constructivist perspective: implications and teaching strategies for science. *School Science and Mathematics, 97(3),* 136-141.

Smith, K. A., & Welliver, P. W. (l990). The development of a science process assessment for forth-grade students. *Journal of Research in Science Teaching, 27(8),* 727-738.